DOI: 10.1111/tgis.12661

END RESEARCH ARTICLE

Transactions in GIS WILEY

A convolutional neural network approach to predict non-permissive environments from moderate-resolution imagery

Seth Goodman[1,2] | Ariel BenYishay[1,3] | Daniel Runfola[2,4]

[1]AidData, Global Research Institute, William & Mary, Williamsburg, VA, USA

[2]Department of Applied Science, William & Mary, Williamsburg, VA, USA

[3]Department of Economics, William & Mary, Williamsburg, VA, USA

[4]Geospatial Evaluation and Observation Lab, Data Science Program, William & Mary, Williamsburg, VA, USA

Correspondence
Seth Goodman, AidData, Global Research Institute, William & Mary, Williamsburg, VA, USA.
Email: smgoodman@wm.edu

Abstract

Convolutional neural networks (CNNs) trained with satellite imagery have been successfully used to generate measures of development indicators, such as poverty, in developing nations. This article explores a CNN-based approach leveraging *Landsat 8* imagery to predict locations of conflict-related deaths. Using Nigeria as a case study, we use the Armed Conflict Location & Event Data (ACLED) dataset to identify locations of conflict events that did or did not result in a death. Imagery for each location is used as an input to train a CNN to distinguish fatal from non-fatal events. Using 2014 imagery, we are able to predict the result of conflict events in the following year (2015) with 80% accuracy. While our approach does not replace the need for causal studies into the drivers of conflict death, it provides a low-cost solution to prediction that requires only publicly available imagery to implement. Findings suggest that the information contained in moderate-resolution imagery can be used to predict the likelihood of a death due to conflict at a given location in Nigeria the following year, and that CNN-based methods of estimating development-related indicators may be effective in applications beyond those explored in the literature.

S.G. conceptualized the project, with input from D.R. and A.B. S.G. wrote code to prepare data and implement machine learning methods, analysis, and other components. D.R. assisted in algorithm and methodology development. This piece was written by S.G. and D.R.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *Transactions in GIS* published by John Wiley & Sons Ltd

1 | INTRODUCTION

Measuring factors surrounding human development around the world is a cornerstone of research and policy within the international development community, as highlighted by targets and indicators of the United Nation's Sustainable Development Goals (Colglazier, 2015). Household surveys and other local data collection methods have historically served as the primary means of gathering data on key indicators associated with human development. Surveys collect data on indicators such as household consumption, assets, health, conflict events and other factors which can be used to target resources, inform policies, and evaluate progress (Dolan et al., 2019; Karlsson, Kim, Joe, & Subramanian, 2020; Sato, 2019).

A major shortcoming of such surveys and measures is the inherent difficulty—and costs—associated with collecting them (Demombynes & Sandefur, 2014; Jerven, 2014). In some cases, survey-based data collection can be impractical due to constraints such as cost and conflict (e.g., Axinn, Ghimire, & Williams, 2012; Coghlan et al., 2006). It is often not possible to collect data with sufficient coverage at the intervals needed. Multiple groups have noted that low-income countries suffer from a substantial data gap due to conducting surveys far less frequently than needed, even for information on basic needs (Chandy & Zhang, 2015).

To overcome data gaps, researchers and policy-makers are increasingly utilizing remotely sensed geospatial data to generate estimates of traditional socioeconomic outcomes and indicators (see Chen & Nordhaus, 2011; Henderson, Storeygard, & Weil, 2012; Noor, Alegana, Gething, Tatem, & Snow, 2008). One of the most prominent examples of this is the use of nighttime lights (NTL) as an indicator of economic activity (Chen & Nordhaus, 2011; Henderson et al., 2012; Noor et al., 2008). While the use of NTL has been shown to correlate with various measures of economic output or wealth in a range of contexts, it is fundamentally limited by the ability to detect low levels of light in poor rural areas (Jean et al., 2016), the relatively coarse resolution of the data,1 and other issues such as light saturation and bloom (Elvidge, Baugh, Zhizhin, & Hsu, 2013).

Recent literature has identified machine learning methods to further improve our ability to detect the geospatial location of impoverished communities. Jean et al. (2016) used a combination of satellite imagery and machine learning to estimate consumption and assets in multiple African countries. This method was based around convolutional neural networks (CNNs), which have been shown to be highly effective with image-based applications (He, Zhang, Ren, & Sun, 2016; Razavian, Azizpour, Sullivan, & Carlsson, 2014). CNN-based methods of estimating poverty using satellite data have now been shown to outperform methods which used NTL alone (Jean et al., 2016). Authors (see, for example, Babenko, Hersh, Newhouse, Ramakrishnan, & Swartz, 2017; Head, Manguin, Tran, & Blumenstock, 2017; Perez et al., 2017) working with this class of methods have begun exploring applications in additional countries, using varying sources of satellite imagery, and addressing potential limitations.

The range of data sources, CNN architectures, and sampling schemes possible for the use of CNN-based approaches to estimation present a number of novel challenges. Additionally, little work has been done to explore the use of these methods to estimate or predict other outcome measures, such as conflict, or assess the effectiveness of these methods when using specific temporal data. This article engages with a subset of these challenges, focusing on the research question whether CNNs can be used to predict likely locations of conflict-related deaths in Nigeria based on satellite imagery. We engage with this as follows. In Section 2 we present background information on conflict in Nigeria, CNNs, and a brief review of the satellite imagery used in this article. In Section 3 we discuss the methods used to prepare the data and implement the CNNs. Section 4 presents the results. Finally, in Section 5, we discuss limitations of the methods and avenues for future work.

2 | BACKGROUND

2.1 | Conflict and instability in Nigeria

Levels of conflict present in Nigeria are the result of a range of complicated political, economic, humanitarian, and ethnic factors—some recent, others rooted in Nigeria's past. Among the most substantial sources of conflict and instability today are the Boko Haram insurgency; violence between farmers and herdsmen often driven into each other's paths by a shifting landscape due to climate change, drought (Akinwotu, 2017), and environmental degradation, along with changing laws; ethnic and religious violence; government and police forces killing civilians; and general corruption (Herbert & Husaini, 2018).

Conflict events in Nigeria, particularly fatal events, are largely tied to the Northern regions where Boko Haram is most active and, according to the Armed Conflict Location & Event Data (ACLED) data set, have been responsible for nearly 11,000 fatalities since 2010 (ACLED, 2019). Another major source of conflict is between herdsmen and farmers. As herdsmen and farmers are forced to move across the land due to environmental factors or grazing laws (International Crisis Group, 2018), conflict in defense of their livelihoods has arisen, with ethnic and religious differences contributing to an already tenuous situation (CFR, 2019; Peace-Insight, 2017). Of the many results of this and other conflicts and instability in Nigeria, and a further factor in future conflict, are internally displaced persons (IDPs) and food insecurity. According to the Famine Early Warning System Network (FEWS Net), there are now over 2.0 million IDPs in the Northeast alone, leading to increased food assistance needs (FEWS, 2019).

Several programs exist to track conflict events in Nigeria with varying mechanisms and scopes. Two of the most commonly used data sets are the ACLED project (Raleigh, Linke, Hegre, & Karlsen, 2010), and the Uppsala Conflict Data Program Georeferenced Event Database (UCDP GED; Sundberg & Melander, 2013). The UCDP data set provides data on organized events of lethal violence, while the ACLED data set provides data on political violence and protest around the world. A key distinguishing factor is that while ACLED can suffer from data quality issues (Eck, 2012; Raleigh et al., 2010), it provides non-lethal conflict event data not available from UCDP (Figure 1). Events included in the UCDP GED are also restricted to those in which at least 25 deaths occurred (Sundberg & Melander, 2013).

Examples of organizations utilizing these data sources or others to track conflict and instability in Nigeria include the Council on Foreign Relations Global Conflict Tracker (CFR, 2019), Peace Insight's conflict profiles (Figure 2) (Peace-Insight, 2017), as well as the Famine Early Warning System Network (FEWS, 2019). Despite the increasing availability and use of georeferenced subnational conflict data, the quality, coverage, and detail of data collected are still fundamentally limited by available reporting and data collection mechanisms (Eck, 2012). In addition to data sets and trackers of historic events, a range of methods have been explored to predict future conflict events, including some utilizing machine learning (Cederman & Weidmann, 2017). However, these methods have been limited in their temporal and spatial precision, as well as their reliability (Bazzi et al., 2019).

Recent works which have shown the viability of using machine learning methods based on convolutional neural networks to estimate poverty (Head et al., 2017; Jean et al., 2016) have been expanded on by researchers to additional applications such as road network quality assessments (Oshri et al., 2018) and crop yield estimates (You, Li, Low, Lobell, & Ermon, 2017) with some success. Adapting these methods to fill in spatial and temporal gaps of conflict, as well as for generating future predictions, could aid the efforts of researchers and actors in the international development community who are currently limited by what data are available.

2.2 | Convolutional neural networks

Convolutional neural networks are a subclass of neural networks which have proven to be remarkably successful in a range of image classification and related computer vision tasks (Albelwi & Mahmood, 2017; Farabet, Couprie,

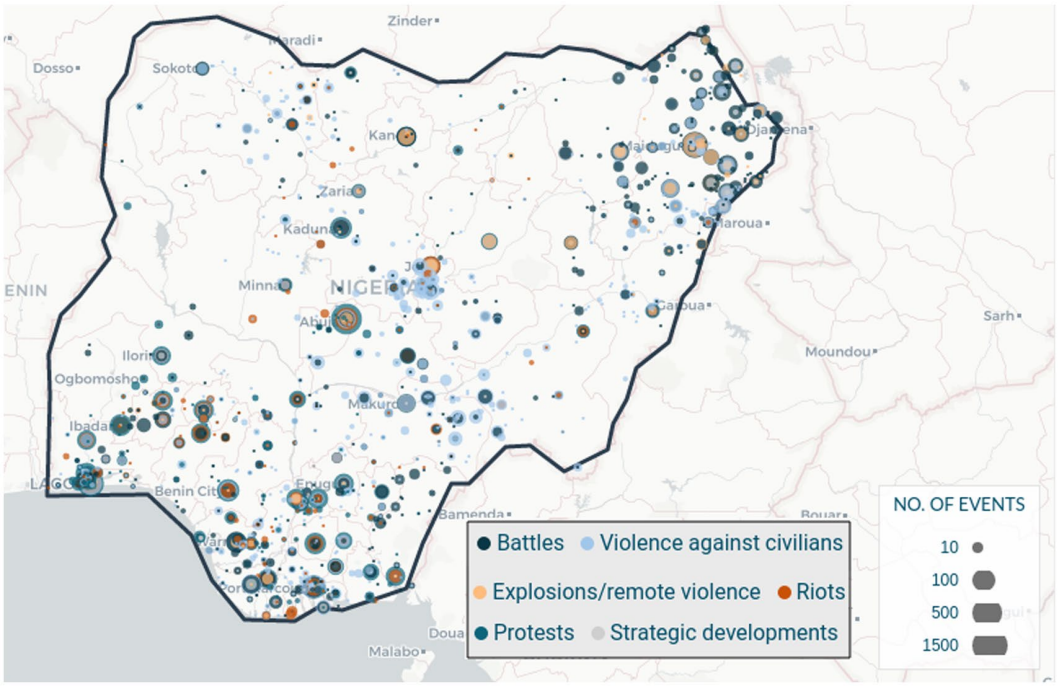

FIGURE 1 All ACLED conflict events in Nigeria (ACLED, 2019; Raleigh et al., 2010)

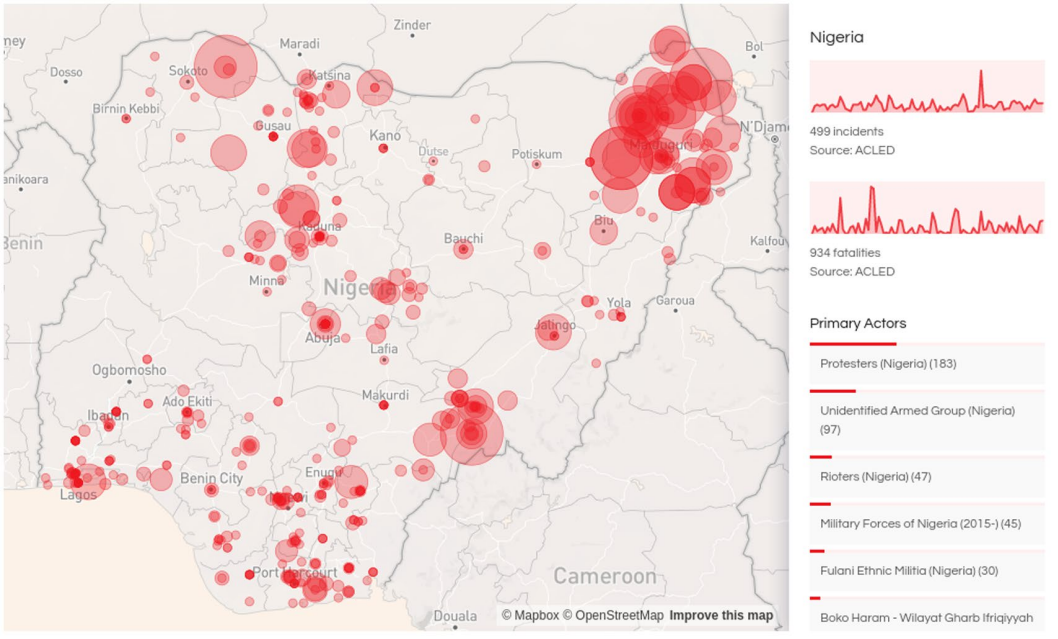

FIGURE 2 Conflict in Nigeria from July through September 2019: approximately 500 incidents resulting in nearly 1,000 deaths (Peace-Insight, 2017; Raleigh et al., 2010)

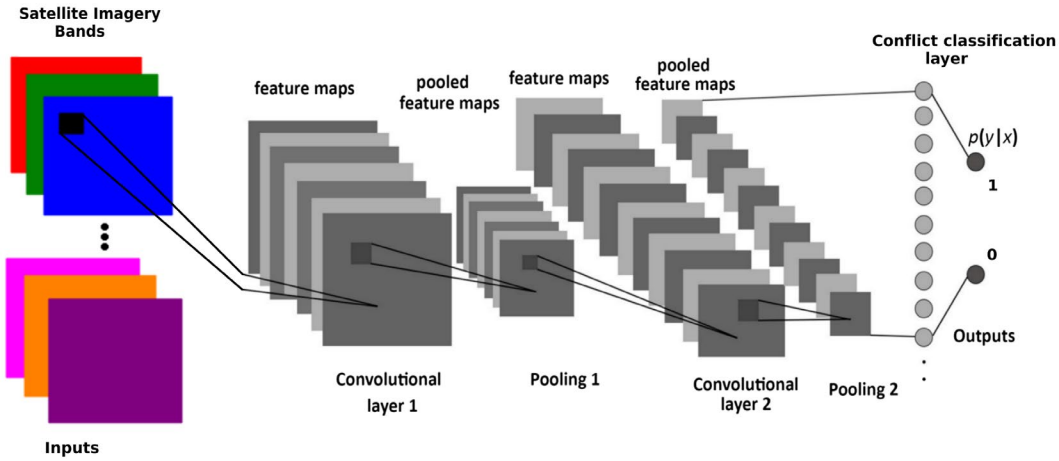

FIGURE 3 Components of a convolutional neural network (Albelwi & Mahmood, 2017)

Najman, & LeCun, 2013; Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014). Inspired by biological structures associated with visual processing, CNNs utilize a unique combination of alternating convolutional and pooling layers. This structure takes advantage of the spatial structure of images (i.e., dimensions) to detect features more efficiently than traditional neural networks (Albelwi & Mahmood, 2017). An example of the structure of CNNs can be seen in Figure 3.

CNNs can be implemented with a variety of architectures which can significantly impact the effectiveness of the CNN, depending on the class of problem being solved (Albelwi & Mahmood, 2017). Classes of CNN architecture that have been developed include LeNet (LeCun, Bottou, Bengio, & Haffner, 1998), AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015), and VGG-Net (Simonyan & Zisserman, 2014). Residual networks, or ResNets, are one of the more recent types of CNNs developed and have been shown to be highly effective in image tasks such as image classification, detection, localization, and segmentation (He et al., 2016). The ability to perform well on traditional image tasks makes ResNets, and CNNs in general, an excellent candidate for utilization in applications based on satellite imagery. To date, CNN-based methods have been used in combination with various sources of satellite imagery to produce estimates of poverty (Perez et al., 2017), crop yields (You et al., 2017), and infrastructure quality (Oshri et al., 2018). While use of such methods is not a guarantee of success (Head et al., 2017), the diversity of promising results from early literature using these methods suggests that other areas and applications may benefit from these methods.

2.3 | Satellite imagery

A key component of the CNN-based approach introduced by Jean et al. (2016) is the source of satellite data used to train the CNNs. Over the past decade a wide range of satellite imagery has become available from private companies such as Digital Global (DigitalGlobe, 2018) and Planet (Boshuizen, Mason, Klupar, & Spanhake, 2014), as well as from free and publicly available sources such as the Landsat (Irons, Dwyer, & Barsi, 2012; USGS, 2018a) and Sentinel (ESA, 2018) programs. Existing work in the development community using CNNs has incorporated satellite data from many satellite imagery sources, including Google Static Maps API (Head et al., 2017; Jean et al., 2016; Xie, Jean, Burke, Lobell, & Ermon, 2015), Planet (Babenko et al., 2017), Digital Globe (*GeoEye-1* and *Quickbird-2*) (Babenko et al., 2017; Engstrom, Hersh, & Newhouse, 2017) *Landsat 7* (Perez et al., 2017), *Landsat 8* (Oshri et al., 2018), and *Sentinel-1* (Oshri et al., 2018). Each satellite has specific sensor capabilities and orbital characteristics associated with the satellite's intended purpose. Factors such as return time and spectral resolution,

which impact the data a satellite collects, are essential considerations when determining an appropriate source of imagery for geospatial applications. Some critical characteristics to consider include:

- *spatial coverage*, the regions of the world which the satellite covers;
- *spatial resolution*, the pixel size, or how fine a feature the satellite can detect;
- *temporal coverage*, the time range over which the satellite has been operating;
- *temporal resolution*, the frequency at which the satellite revisits a location; and
- *spectral resolution*, the number and range of bands (ranges of light) that the sensor can pick up.

When using satellite imagery to estimate metrics of human development, the above factors will influence the accuracy and suitability of different CNN architectures. Spatial resolution impacts both the ability of the CNN to detect appropriate features, and will ultimately impact the resolution or precision at which the final estimates can be generated (Blaschke, 2010). For example, imagery with a resolution of 1 km would have limited use if the goal was to detect the roof material of individual houses in rural area. Finally, satellite platforms can contain a range of sensors with the ability to detect varying spectral characteristics. For example, if an application depends on the detection of features only visible using infrared imagery, then a satellite which can only capture visible wavelengths (e.g., red, green, and blue bands) would not be suitable.

2.3.1 | Landsat imagery

The Landsat program includes a temporally continuous series of satellites that offer a combination of spatial, temporal, and spectral resolution and coverage that make it a powerful tool for regularly monitoring changes of relevance to policy-makers. *Landsat 8* is the most recent active satellite in the Landsat program since its launch in 2013, and its data are public. *Landsat 8* captures data for the entire planet, revisiting each location at 16-day intervals. The sensors on board *Landsat 8* acquire imagery at a 30 m resolution in the visible, near infrared, and shortwave infrared range using the Operation Land Imager sensor, as well as thermal bands from the Thermal Infrared Sensor. These bands are detailed in Table 1.

Landsat 7, the predecessor to *Landsat 8*, remains active until 2021 (USGS, 2017a) and offers significant temporal coverage from 1999 to the present. *Landsat 7* also shares many of the attributes that make *Landsat 8* useful (USGS, 2018b). When combined with broad temporal coverage, these attributes have made *Landsat 7* useful

TABLE 1 *Landsat 8* operational land imager and thermal infrared sensor bands

Bands	Wavelength (μm)	Resolution (m)
Band 1 – Coastal aerosol	0.433–0.453	30
Band 2 – Blue	0.450–0.515	30
Band 3 – Green	0.525–0.600	30
Band 4 – Red	0.630–0.680	30
Band 5 – Near infrared	0.845–0.885	30
Band 6 – Shortwave infrared 1	1.560–1.660	30
Band 7 – Shortwave infrared 2	2.100–2.300	30
Band 8 – Panchromatic	0.500–0.680	15
Band 9 – Cirrus	1.360–1.390	30
Band 10 – Thermal infrared 1	10.60–11.20	30[a]
Band 11 – Thermal infrared 2	11.50–12.50	30[a]

[a]Collected at 100 m but resampled to 30 m to match OLI bands.

for a wide range of historic and contemporary applications focused on regions around the world (Demirkesen, Evrendilek, Berberoglu, & Kilic, 2007; Jensen & Cowen, 1999; USGS, 2018a; Yang, Huang, Homer, Wylie, & Coan, 2003). A major drawback to using *Landsat 7* is the scan line corrector (SLC) failure which occurred aboard *Landsat 7* in 2003 (USGS, 2018d).

The SLC failure impacted the satellite's ability to correct for forward movement, leaving gaps in the resulting images, as seen in Figure 4. A range of methods exist to fill in the gaps created by the SLC failure (Maxwell, 2004; Scaramuzza, Micijevic, & Chander, 2004; Zeng, Shen, & Zhang, 2013; Zhu, Liu, & Chen, 2012) and allow *Landsat 7* data to continue to be used in some applications. For applications which do not require a historic record, *Landsat 8* offers very similar data without the need for additional gap filling. In addition to the availability of historic data from *Landsat 7* and earlier satellites in the program (USGS, 2019), long-term support for the Landsat series is already planned, with the launch of *Landsat 9* scheduled for late 2020.

Several alternatives to Landsat exist, and often offer improvements along certain dimensions. For example, significantly finer spatial resolution imagery is available from Digital Globe or Planet (sub-meter resolution). However, these products are not freely available, and do not have the spectral or temporal coverage available from Landsat (Boshuizen et al., 2014; DigitalGlobe, 2018). One of the most competitive alternatives to Landsat is the Sentinel program. The *Sentinel-1* and *Sentinel-2* satellites have offered publicly available data, since 2014 and 2015 respectively. *Sentinel-2* has a resolution of 10–20 m and is only lacking thermal bands when compared to *Landsat 8* (Drusch et al., 2012), while *Sentinel-1* utilizes a C-band synthetic-aperture radar instrument (Geudtner, Torres, Snoeij, Davidson, & Rommen, 2014) for collecting data. Initial work has shown success in utilizing *Sentinel-1* data in satellite-imagery-based machine learning applications (Oshri et al., 2018), and the potential for exploring the use of *Sentinel-2* will be discussed in Section 5.

Despite the relatively coarse resolution of Landsat imagery, existing work has shown that when used with CNN-based approaches for predicting poverty a resolution of 30 m can still be effective (Perez et al., 2017). This previous success in similar applications, combined with the aforementioned favorable characteristics of *Landsat 8*, make this product a viable satellite imagery candidate for exploring the effectiveness of CNN-based methods to accurately predict conflict event fatalities.

FIGURE 4 Example of SLC-off imagery (Scaramuzza et al., 2004)

3 | METHODS

3.1 | Data preparation

The primary source of input data for this application is daytime satellite imagery from *Landsat 8*. Imagery was requested and downloaded using the U.S. Geological Survey's freely available EarthExplorer (USGS, 2017b) and Bulk Download tools (USGS, 2017c). High-quality imagery suitable for time-series analysis2 was downloaded for Nigeria for all of 2014. Sixty scenes3 of imagery are acquired from each full revisit cycle to achieve complete coverage of Nigeria. Due to variability in the quality of the raw data collected by the sensors not all scenes captured at each revisit are suitable for use in analysis. Each individual scene contains ten bands of data at 30 m resolution, corresponding to the wavelengths detailed in Table 1.

The analysis in this article leverages yearly time-steps, necessitating additional processing. Yearly composites are created from individual scenes by taking the mean pixel values within each imagery scene at every available time-step for the year. Prior to aggregation, each individual scene is masked for pixels associated with cloud cover. These pixels are defined by pixel quality assurance data which accompany each scene and are provided by USGS (2018c). All pixels with values indicating high or medium confidence of cloud coverage are masked during processing. Once all available scenes captured over the year for individual locations are masked and aggregated, the resulting set of scenes is then combined into a single mosaic of images covering the entire country.

Conflict data were acquired from the ACLED database (Raleigh et al., 2010). The location of events in ACLED is defined by longitude and latitude, and is used to define the satellite imagery used for training and validation data. ACLED data also provide the number of known fatalities for each conflict event, which is used to create a binary classification of either no fatalities (0) or fatalities (1) for each event. The binary class is then used to label the associated imagery.

According to ACLED, there were 1,673 conflict events in Nigeria in 2015. Of these, 680 resulted in fatalities and 993 did not. The locations associated with these events were used to generate samples for training (85% of the data) and validation (15%). The number of samples for each of the binary classes was then reduced based on the smaller class size to create a balanced set of samples for each class. Each sample consists of a 224 × 224 pixel image for each band extracted from the 2014 *Landsat 8* mosaic generated earlier. The dimensions of these samples correspond to input dimensions used for the class of CNN known as ResNets, discussed in Section 3.2.

In addition to utilizing each conflict location for a single sample, two additional sampling methods are implemented for comparison. The primary comparison method will attempt to determine if nearby spatial features in the imagery improve the model's performance. For each original sample location, this method generates nine additional samples based on points distributed in a regular grid around the original location within approximately 1,000 m. To ensure that the increased sample size is not responsible for improved accuracy, as opposed to the content of the additional samples, one additional comparison method will be tested. This secondary comparison method involves duplicating each original sample nine times and will serve as a balanced sample size comparison for the previous method.4

3.2 | CNN implementation

The CNNs tested in this article were implemented using PyTorch, an open source Python package, and run using Nvidia Tesla GPUs.5 The specific class of CNN implemented is a residual network (ResNet), which has been developed with varying architectures that are primarily defined by the depth of the network (e.g., ResNet-18 has 18 layers). Using PyTorch, ResNets can be initialized using pre-trained weights based on training performed using ImageNet. ImageNet is a collection of approximately 14 million images from over 20,000 categories, including cars, dogs, and cats (Deng et al., 2009). Utilizing pre-trained networks enables a network to leverage information

learned by training on a far larger data set, which can then be incorporated and refined by an application using a smaller, limited, data set. This process is known as transfer learning.

The fundamental principle of transfer learning is that training on separate yet similar data set will allow the network to learn basic, generalizable features that are useful for any image classification task. Existing work has shown that using transfer learning to initialize a network, even when the original training task is extremely different, can still outperform traditional weight initialization techniques (Pan & Yang, 2010; Weiss, Khoshgoftaar, & Wang, 2016; Yosinski, Clune, Bengio, & Lipson, 2014). Approaches to machine learning based on transfer learning have grown in popularity as they have enabled machine learning to be used in applications with limited data that would not previously have been possible (Hoo-Chang et al., 2016; Jean et al., 2016; Karpathy et al., 2014; Razavian et al., 2014).

Once the network is initialized with the pre-trained weights, the data for the new task can be used to update the network. A common procedure for updating the network involves "freezing" the earlier layers, where basic features are learned, and only fine-tuning deeper layers which will impact more complicated, application-specific, features (Yosinski et al., 2014). Then, the final fully connected layer which is responsible for performing the actual classification—based on the output from earlier layers—is modified and trained to reflect the current task.

A critical difference between the images from ImageNet used for pre-training and satellite imagery is the inclusion of additional bands of data. Traditional images contain only red, green, and blue (RGB) bands to reflect the visible wavelengths of imagery, whereas *Landsat 8* includes a total of 10 bands that cover a broader spectrum of imagery. Due to this difference, the CNN must be adapted to consider the extra bands of data. To utilize the additional bands of data available with *Landsat 8*, a custom data loader was created6 and the initial convolutional layer of the CNN modified based on the number of input bands. Additionally, each new band must have weights initialized. Following the implementation strategy of previous work (Perez et al., 2017), the average of the pre-trained weights for the RGB bands was used to initialize the new bands.

3.3 | Optimization

A range of different CNN architectures and parameters are tested to determine the optimal combination for accurately predicting conflict fatalities. A grid search is used to iterate over a defined set of parameters for each variable associated with the CNN. The variables explored include:

- *network architecture*, the different versions of ResNet based on varying network depth (number of layers);
- *learning rate*, a factor influencing the impact of training data on the network, or how fast the network learns from data;
- *momentum*, a factor which helps to overcome issues associated with local minima;
- *step size*, the number of iterations at which learning rate is decayed;
- *gamma*, the amount the learning rate is decayed at a specified step size; and
- *optimization function*, the type of optimization algorithm used.

For each unique combination of these variables, the accuracy of a CNN is estimated based on the percentage of ACLED locations it successfully classifies as either a fatal (1) or non-fatal (0) conflict event. This is done by progressively training the CNN over multiple iterations, or epochs, for each set of parameters. Each epoch consists of passing the entirety of the training data through the CNN, allowing it to learn, and then using the validation data to assess the accuracy of the CNN for that epoch. Each subsequent epoch builds on the learned behavior of the previous epochs, assessing accuracy until the defined number of epochs has been reached.7

The accuracy of a given CNN, consisting of one set of parameters from the grid search, is defined as the maximum epoch accuracy achieved. As each epoch is progressively run, it is not always the case that the learned information improves the network's accuracy. In order to determine the most effective network, the state of the

network and its accuracy are recorded after each epoch. Once all epochs have been completed, the network state with the best accuracy is saved.

4 | RESULTS

To assess the ability of a convolutional neural network to predict conflict fatalities, pre-trained ResNet CNNs were implemented in PyTorch as detailed in Section 3.2. Multi-spectral *Landsat 8* imagery for 2014 and ACLED conflict event data for 2015 were prepared based on the methods in Section 3.1. The subsequent year (2015) of conflict data was selected in order determine the potential effectiveness of these methods to predict future conflict fatalities based on contemporary satellite imagery. Restricting data within temporal bounds is a key feature that distinguishes this application from earlier work (e.g., Jean et al., 2016) which will provide insights into the feasibility of this approach for time series predictions. Using the locations and fatality count of conflict events, training and validation data sets were labeled using a binary of whether conflict events did or did not have any fatalities associated with them. The set of samples for each binary class in the training and validation data were balanced to ensure no bias during training. Two sets of tests were performed which were designed to identify the most effective sampling method for training data, as well as identify the optimal combination from a range of network parameters.

The three sampling strategies tested were based on a combination of machine learning and spatial theory, and determined the satellite image(s) that were input into the network for each observation. Method 1 (no fill) used a single 244 × 244 pixel window (for each band) around the latitude–longitude location of the conflict event, and is designed to serve as a baseline for comparison. Method 2 (grid fill) used nine equally spaced samples in a grid around the original sample latitude and longitude, to test the theory that an expanded geographic region of information could provide improved predictions. Method 3 (duplicates) duplicated the original location nine times, and served as a comparison with a similar sample size to method 2 in order to avoid bias due to sample size.

During the first set of tests, each sampling method was tested using 64 different parameter combinations over a training and validation cycle of 30 epochs, detailed in Table 2. Based on these tests, both methods which expanded the size of the training data set resulted in a clear improvement in accuracy over the baseline method, as seen in Figure 5. Methods 2 and 3 have similar minima during this initial set of tests, but method 3 achieved higher median and maximum accuracy values. Individual parameter variations across the sampling methods generally did not show significant performance trends (Figure 6)—the only notable finding was slightly better performance when using the smaller learning rate.

The second set of tests expanded the range of parameters explored (Table 2) while using method 3, for a total of 432 parameter combinations. Method 3 was chosen as the focus for the second set of tests due to its strong performance in the first set of tests. For the second set of tests, the number of epochs run was increased to 60. More variation was found across the expanded batch of tests, with a minimum accuracy of 59.8% and a maximum of 80.9%. However, no clear patterns emerged among the parameter combinations aside from those associated with learning rate. Smaller learning rates outperformed larger learning rates as they did in the initial batch of

TABLE 2 CNN parameters

Parameter	Batch 1 values	Batch 2 values
Network architecture	ResNet18, ResNet50	ResNet18, ResNet50
Learning rate	0.01, 0.001	0.1, 0.01, 0.001, 0.0001
Momentum	0.95, 0.99	0.95, 0.97, 0.99
Step size	5, 10	5, 10, 15
Gamma	0.25, 0.75	0.25, 0.5, 0.75
Optimization function	SGD, Adam	SGD, Adam

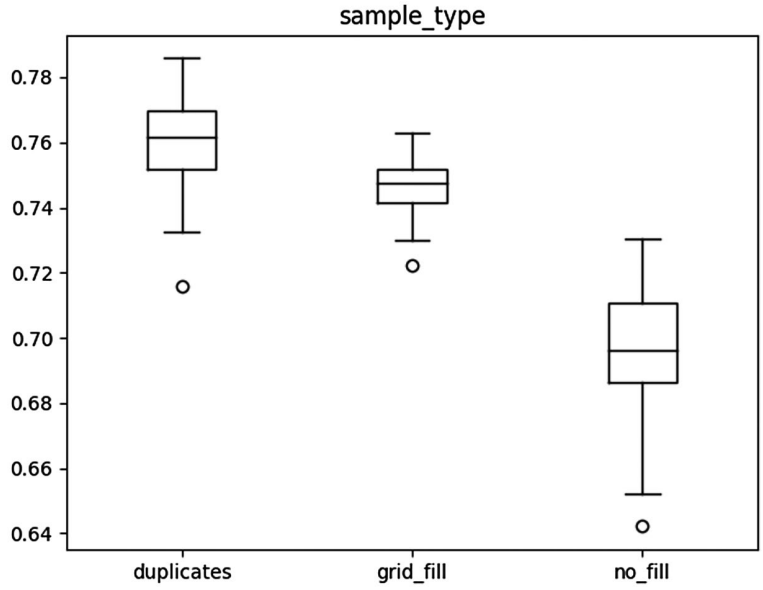

FIGURE 5 Boxplot of sample type

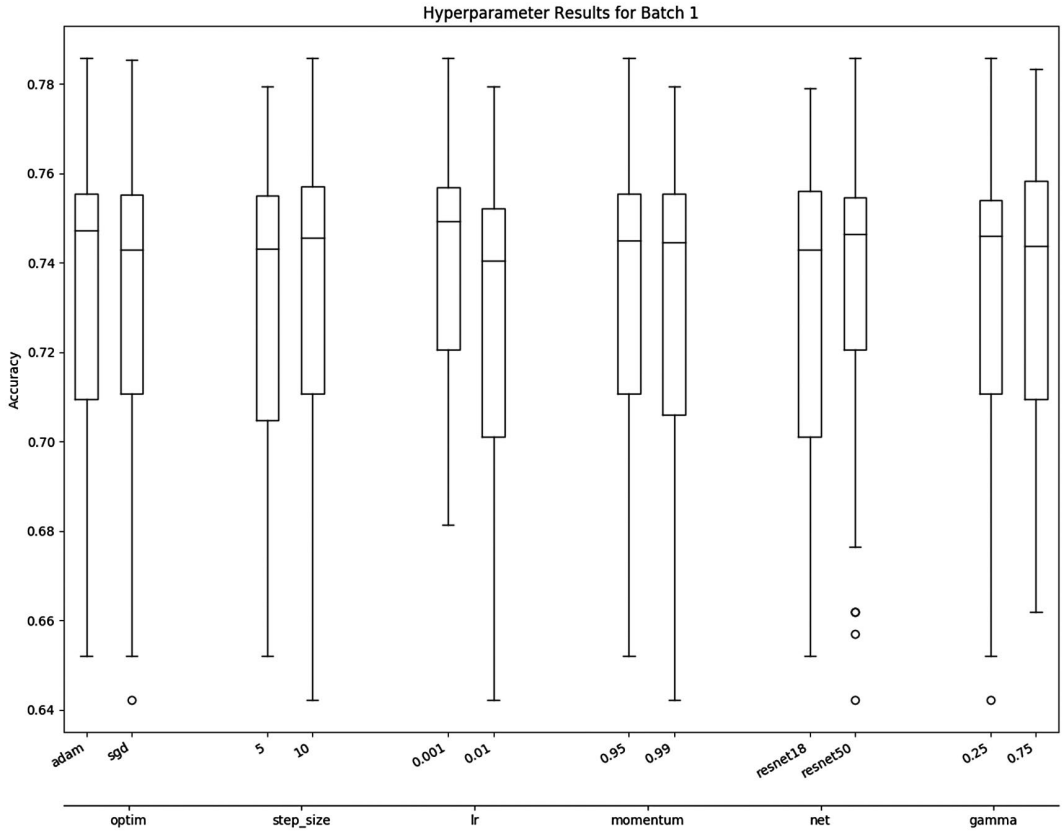

FIGURE 6 Boxplot of parameters, first set of tests

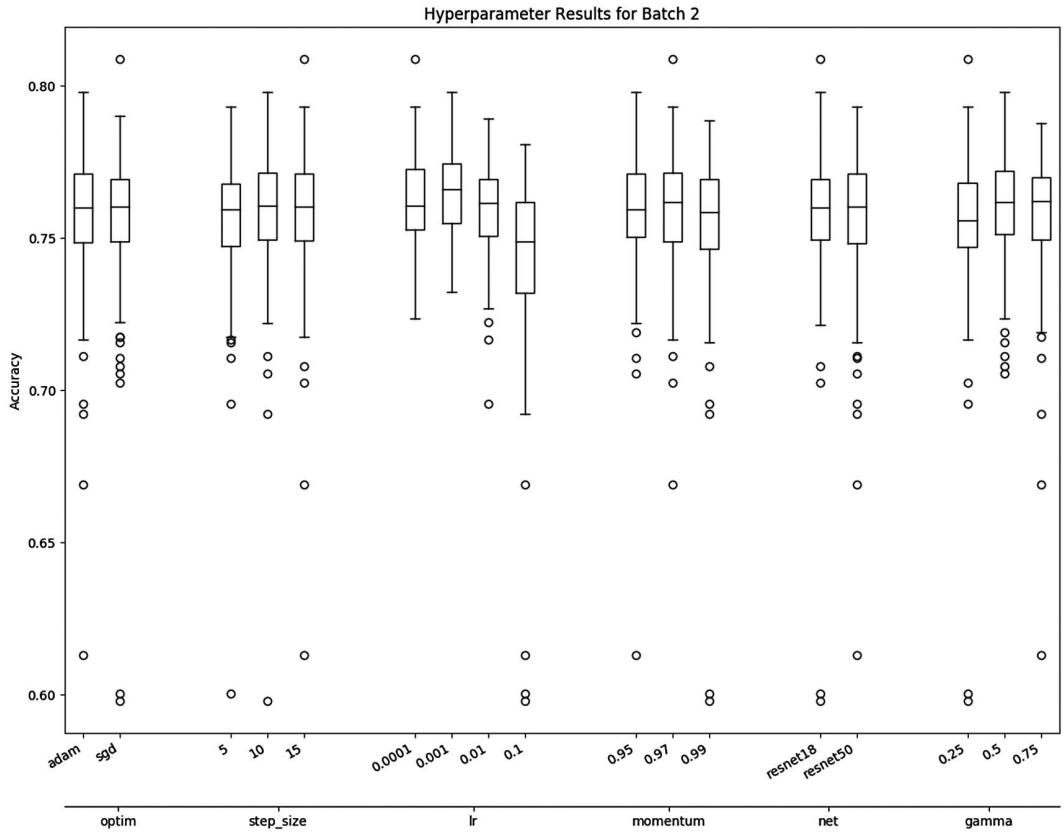

FIGURE 7 Boxplot of parameters, second set of tests

tests. The best overall accuracy was achieved using a learning rate of 0.0001, but overall trends showed no significant performance improvement over a learning rate of 0.001. As expected given the balanced class sizes during training, predictions for neither class skewed the overall accuracy. Using the best-performing parameters (overall accuracy of 80.9%), the accuracy for predicting conflict events with fatalities (true positives) was 79.4%, and the accuracy for predicting conflict events with no fatalities (true negatives) was 82.4%.

Overall, the grid search of parameters conducted across the two batches of tests showed minimal impact from parameter adjustments within commonly accepted ranges (see Figure 7). The only notable exception to this was large (i.e. multiple orders of magnitude) changes in learning rate. The sampling method and the resulting data used for training proved to be a much more significant factor in the overall ability of a network to accurately identify the likelihood of future conflict in a region based on satellite imagery. Additionally, these results were compared to CNNs which were trained using randomly labeled data. The accuracy of random data networks after multiple iterations was found to be significantly lower (maximum of approximately 50%) when compared to the true network (maximum of over 80%).

5 | DISCUSSION AND CONCLUSIONS

5.1 | Convolutional neural network findings

The work presented in this article provides evidence that CNNs applied to moderate-resolution satellite imagery can be used to predict the likelihood of conflict-related deaths in Nigeria. We specifically illustrate that if a conflict

event occurs in 2015, multi-spectral imagery of the physical features on the ground at that location in 2014 can predict with 80% accuracy whether or not the conflict event will result in any fatalities. This finding illustrates the value of spatial information for prediction of development indicators, especially when processed with a CNN-based approach. While the use of the methods presented in this article for predicting non-permissive environments and related policy applications requires additional research, our results show the potential for CNN-based methods in applications beyond those explored in the existing literature.

One of the most notable findings of this work is the limited impact of varying network parameters when compared to the impact of modifying sampling schemes for training data. Given the impact of the sampling method on overall network performance, expanding the scope of the tests comparing these methods may lead to significant improvements. Utilizing duplicates of training data was identified as the most effective sampling method, yet this duplication could simply offer similar performance gains as a proportional increase in the number of epochs during training. Comparing duplication at varying levels with proportional increases in the number of epochs could provide insight into the impact each of the strategies has on performance. Further, while the grid-based approach to creating additional training data samples underperformed relative to simple duplication of training data baseline, additional work which explores CNN accuracy when varying the distance from the original sample location to generate the grid, as well as the number of additional points in the grid, could prove more effective and provide additional insight into the spatial correlation of satellite imagery and conflict fatalities.

Considering the significance of training data, comparing the currently used data from *Landsat 8* and ACLED with other sources of satellite imagery and conflict data would also be valuable. Of other available sources of satellite imagery, discussed in Section 2.3, the *Sentinel-2* platform offers the most compelling case for comparison. *Sentinel-2* is publicly available, has similar spectral coverage to *Landsat 8*, and has finer resolution. Other sources of conflict and event data such as the Social Conflict Analysis Database (SCAD; Salehyan et al., 2012) and the Global Database of Events, Language, and Tone (GDELT Leetaru & Schrodt, 2013) could be used as alternatives or supplements to ACLED data.

Another limiting factor of our approach is that the training data were restricted to conflict events and classified by whether each event did or did not have fatalities associated with it. An alternative approach, which could be achieved using random locations or non-conflict event locations, would be to compare conflict events with control locations at which no conflict occurred. Additional methods for testing the robustness of our approach to predicting conflict include expanding the temporal and spatial coverage using historical data. The work presented in this article used only 2014 imagery and 2015 conflict data, while data for both exist through 2019. Additionally, both imagery and ACLED data exist for other countries which could be explored independently or as part of a comprehensive model. Testing across countries would also contribute to exploring the impact of spatial scale of training on network performance. Models could also be tested which restrict data to subnational regions by way of comparison.

5.2 | Core limitations of a CNN-based predictive approach

5.2.1 | Scope and extent

This article introduces a purely predictive CNN-based approach for identifying whether a conflict event is likely to result in a death: we make no causal inferential claims, nor does the current CNN approach support the examination of causality. The causes, factors, and defining characteristics of conflict death will frequently be very different, and be associated with different features found in satellite imagery, when compared with another conflict. As such, CNN-based predictive models trained solely on data from Nigeria may not perform effectively in other countries without including those countries in the training.

Given that geographic extent could reasonably be a limiting factor in effectiveness of this approach, we explored the spatial distribution of both classes of conflict events, as well as the spatial distribution of locations

which were accurately (and inaccurately) predicted in Nigeria. In all cases, no clear pattern was found, indicating it is unlikely that our model was skewed toward only predicting accurately in certain portions of the country with homogeneous spatial features. While this does not indicate that this specific model could be applied to other countries, it is indicative of good overall performance across varying geographic contexts for conflict in Nigeria.

5.2.2 | Intractability

As we consider that the landscape and features associated with conflict may vary across geography, it is important to recognize that the CNN approach detailed here does not attempt to define the specific features being detected. Thus, we are limited to general statements about the type of features likely to be detected given the moderate 30 m resolution of the imagery. Small features such as broken windows or burned cars will not be detectable with 30 m resolution data. Rather, the CNN is only capable of detecting large landscape trends such as urban areas, draught-stricken areas, or other features which may be well-established correlates of conflict from the literature. While CNN methods may have the potential to help identify novel indicators of conflict, identifying what specifically the CNN is detecting is beyond the scope of this article and would be a valuable focus of future research.

More broadly, despite the promising results presented here, these methods are not intended to supplant existing data sources or methods used for assessing conflict. While these methods may serve as a useful supplement within a broader ecosystem of conflict tools and methodologies, the goal at this stage is to serve as a spur for additional work on the applications of CNN approaches to predicting conflict as well as other development indicators.

5.3 | Conclusions

This paper sought to answer the research question whether convolutional neural networks can be used to predict likely locations of conflict-related deaths in Nigeria based on satellite imagery. The approach presented shows promising results, achieving up to 80% accuracy in binary classification of death versus no death conflict events in the year after model calibration. Despite the significance of this result, we highlight the many limitations of the CNN-based approach to conflict prediction accomplished to date, the most predominant being the lack of insight into what physical features in imagery the CNN identifies as related to conflict deaths. Future research should engage with many of these challenges, including how to identify the driving image features detected by a convolutional neural network, and how to better understand the spatial scopes across which a given model calibration might be most effective. Despite these limitations, the public availability of the data and tools used and the high temporal frequency at which satellite imagery is available allow for a broad range of potential applications and adaptation to various use cases. Examples include the use of predictive maps, generated using these methods, as supplemental aids in decision-making processes for security specialists implementing aid projects in developing countries.

ACKNOWLEDGEMENTS

This study was funded by a cooperative agreement (AID-OAA-A-12-00096) between USAID's Global Development Lab and AidData at the College of William & Mary under the Higher Education Solutions Network (HESN) Program. The authors acknowledge William & Mary Research Computing for providing computational resources and technical support that have contributed to the results reported within this article.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ORCID

Seth Goodman http://orcid.org/0000-0002-5138-9349

ENDNOTES

1 Resolution of 1 km for NTL from the Defense Meteorological Satellite Program (DMSP) which has time series data available for 1992–2013 (NOAA Earth Observation Group, 2017), and 500 m for the Visible Infrared Imaging Radiometer Suite (VIIRS), available from 2012 to the present (Elvidge, Baugh, Zhizhin, Hsu, & Ghosh, 2017).

2 Imagery scenes with the highest data quality, and suitable for time-series analysis, are labeled as Tier 1. This includes precision and terrain correct data that are intercalibrated across Landsat sensors (USGS, 2018c).

3 Each scene is approximately 185 × 180 km and has varying overlap with neighboring scenes.

4 Note that training and validation data were split prior to generating additional samples. Any additional samples generated exist solely in the same training or validation group as the original sample they are based on.

5 All CNNs described in this article are run on either an Nvidia Tesla V100 or P100 GPU, with 16 GB of memory. Software versions include CUDA 9.1, PyTorch 0.4.0, and Python 2.7.13.

6 A custom PyTorch data set class was created which takes in a list of coordinates defining training and validation data coordinates, along with their labels, and extracts a 224 × 224 × 10 pixel image from the country mosaics of Landsat imagery which can be then passed to the CNN. The standard dimensions for a single band of an image from ImageNet are 224 × 224. A complete RGB image would be 224 × 224 × 3.

7 This step reflects a key component of any machine learning application—distinct training and validation data sets. If the same data were used for both training and validation, accuracy would likely consistently improve with additional epochs, yet it would result in an CNN that has been overfitted, and would likely perform poorly when using other data.

REFERENCES

ACLED. (2019). *ACLED data dashboard*. Retrieved from https://www.acleddata.com/dashboard/#566

Akinwotu, E. (2017). Drought worsens deadly battle between Fulani herdsmen and farmers in Nigeria. *Guardian*, January 3. Retrieved from https://www.theguardian.com/global-development/2017/jan/03/drought-worsens-deadly-conflict-between-fulani-herdsmen-nigeria-farmers

Albelwi, S., & Mahmood, A. (2017). A framework for designing the architectures of deep convolutional neural networks. *Entropy*, *19*(6), 242.

Axinn, W. G., Ghimire, D., & Williams, N. E. (2012). Collecting survey data during armed conflict. *Journal of Official Statistics*, *28*(2), 153.

Babenko, B., Hersh, J., Newhouse, D., Ramakrishnan, A., & Swartz, T. (2017). Poverty mapping using convolutional neural networks trained on high and medium resolution satellite images, with an application in Mexico. Preprint, *arXiv*:1711.06323

Bazzi, S., Blair, R. A., Blattman, C., Dube, O., Gudgeon, M., & Peck, R. M. (2019). *The promise and pitfalls of conflict prediction: Evidence from Colombia and Indonesia* (NBER Working Paper No. 25980). Cambridge, MA: National Bureau of Economic Research.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry & Remote Sensing*, *65*(1), 2–16.

Boshuizen, C., Mason, J., Klupar, P., & Spanhake, S. (2014). Results from the Planet Labs flock constellation. In *Proceedings of the 28th AIAA / USU Small Satellite Conference*, Logan, UT (pp. 1–8). Reston, VA: AAAI.

Cederman, L. E., & Weidmann, N. B. (2017). Predicting armed conflict: Time to adjust our expectations? *Science*, *355*(6324), 474–476.

CFR. 2019. *CFR Nigeria security tracker.* Retrieved from https://www.cfr.org/nigeria/nigeria-security-tracker/p29483

Chandy, L., & Zhang, C. (2015). *Stuffing data gaps with dollars: What will it cost to close the data deficit in poor countries?* Retrieved from https://www.brookings.edu/opinions/stuffing-data-gaps-with-dollars-what-will-it-cost-to-close-the-data-deficit-in-poor-countries/

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences of the USA*, *108*(21), 8589–8594.

Coghlan, B., Brennan, R. J., Ngoy, P., Dofara, D., Otto, B., Clements, M., & Stewart, T. (2006). Mortality in the Democratic Republic of Congo: A nationwide survey. *Lancet*, *367*(9504), 44–51.

Colglazier, W. (2015). Sustainable development agenda: 2030. *Science*, *349*, 1048–1050.

Demirkesen, A. C., Evrendilek, F., Berberoglu, S., & Kilic, S. (2007). Coastal flood risk analysis using Landsat-7 ETM+ imagery and SRTM DEM: A case study of Izmir, Turkey. *Environmental Monitoring and Assessment*, *131*(1), 293–300.

Demombynes, G., & Sandefur, J. (2014). *Costing a data revolution*. Retrieved from https://www.cgdev.org/publication/costing-data-revolution

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, FL (pp. 248–255). Piscataway, NJ: IEEE.

DigitalGlobe. (2018). *DigitalGlobe*. Retrieved from https://www.digitalglobe.com/

Dolan, C. B., BenYishay, A., Grépin, K. A., Tanner, J. C., Kimmel, A. D., Wheeler, D. C., & McCord, G. C. (2019). The impact of an insecticide treated bednet campaign on all-cause child mortality: A geospatial impact evaluation from the Democratic Republic of Congo. *PLoS ONE*, *14*(2), e0212890.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., … Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, *120*, 25–36.

Eck, K. (2012). In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Cooperation & Conflict*, *47*(1), 124–141.

Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., & Ghosh, T. (2017). VIIRS night-time lights. *International Journal of Remote Sensing*, *38*(21), 5860–5879.

Elvidge, C. D., Baugh, K. E., Zhizhin, M., & Hsu, F.-C. (2013). Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network*, *35*, 62–69.

Engstrom, R., Hersh, J. S., & Newhouse, D. L. (2017). *Poverty from space: Using high-resolution satellite imagery for estimating economic well-being*. Retrieved from http://documents.worldbank.org/curated/en/610771513691888412/Poverty-from-space-using-high-resolution-satellite-imagery-for-estimating-economic-well-being

ESA. (2018). *Sentinel*. Retrieved from https://sentinel.esa.int/web/sentinel/home

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1915–1929.

FEWS. (2019). *Nigeria famine early warning systems network*. Retrieved from https://fews.net/west-africa/nigeria/key-message-update/september-2019-0

Geudtner, D., Torres, R., Snoeij, P., Davidson, M., & Rommen, B. (2014). Sentinel-1 system capabilities and applications. In *Proceedings of the 2014 International Geoscience and Remote Sensing Symposium*, Québec City, Canada (pp. 1457–1460). Piscataway, NJ: IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NY (pp. 770–778). Piscataway, NJ: IEEE.

Head, A., Manguin, M., Tran, N., & Blumenstock, J. E. (2017). Can human development be measured with satellite imagery? In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*, Lahore, Pakistan (pp. 8:1–8:11). New York, NY: ACM.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, *102*(2), 994–1028.

Herbert, S., & Husaini, S. (2018). *Conflict, instability and resilience in Nigeria*. Retrieved from https://gsdrc.org/publications/conflict-instability-and-resilience-in-nigeria/

Hoo-Chang, S., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., … Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285.

International Crisis Group. (2018). *Stopping Nigeria's spiralling farmer-herder violence*. Retrieved from https://www.crisisgroup.org/africa/west-africa/nigeria/262-stopping-nigerias-spiralling-farmer-herder-violence

Irons, J. R., Dwyer, J. L., & Barsi, J. A. (2012). The next Landsat satellite: The Landsat data continuity mission. *Remote Sensing of Environment*, *122*, 11–21.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794.

Jensen, J. R., & Cowen, D. C. (1999). Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering & Remote Sensing*, *65*, 611–622.

Jerven, M. (2014). *Benefits and costs of the data for development targets for the post-2015 development agenda* (Working Paper). Copenhagen, Denmark: Copenhagen Consensus Center.

Karlsson, O., Kim, R., Joe, W., & Subramanian, S. V. (2020). The relationship of household assets and amenities with child health outcomes: An exploratory cross-sectional study in India, 2015–2016. *SSM - Population Health*, *10*, 100513.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the 27th IEEE conference on Computer Vision and Pattern Recognition*, Columbus, OH (pp. 1725–1732). Piscataway, NJ: IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *60*(6), 1097–1105.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Leetaru, K., & Schrodt, P. A. (2013). Gdelt: Global data on events, location, and tone. In *Proceedings of the 2013 ISA Annual Convention*, San Francisco, CA (pp. 1979–2012). Storrs, CT: ISA.

Maxwell, S. (2004). Filling Landsat ETM+ SLC-off gaps using a segmentation model approach. *Photogrammetric Engineering & Remote Sensing*, *70*(10), 1109–1112.

NOAA Earth Observation Group. (2017). *DMSP-OLS nighttime lights*. Retrieved from https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

Noor, A. M., Alegana, V. A., Gething, P. W., Tatem, A. J., & Snow, R. W. (2008). Using remotely sensed night-time light as a proxy for poverty in Africa. *Population Health Metrics*, *6*(1), 5.

Oshri, B., Hu, A., Adelson, P., Chen, X., Dupas, P., Weinstein, J., Burke, M., … Ermon, S. (2018). Infrastructure quality assessment in Africa using satellite imagery and deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, London, UK (pp. 616–625). New York, NY: ACM.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge & Data Engineering*, *22*, 1345–1359.

Peace-Insight. (2017). *Nigeria: Conflict profile*. Retrieved from https://www.peaceinsight.org/conflicts/nigeria/

Perez, A., Yeh, C., Azzari, G., Burke, M., Lobell, D. B., & Ermon, S. (2017). Poverty prediction with public Landsat 7 satellite imagery and machine learning.Preprint, *arXiv*:1711.03654.

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An armed conflict location and event dataset: Special data feature. *Journal of Peace Research*, *47*(5), 651–660.

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. In *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, OH (pp. 512–519). Piscataway, NJ: IEEE.

Salehyan, I., Hendrix, C. S., Hamner, J., Case, C., Linebarger, C., Stull, E., & Williams, J. (2012). Social conflict in Africa: A new database. *International Interactions*, *38*(4), 503–511.

Sato, R. (2019). Effect of armed conflict on vaccination: Evidence from the Boko Haram insurgency in northeastern Nigeria. *Conflict & Health*, *13*, 49.

Scaramuzza, P., Micijevic, E., & Chander, G. (2004). *SLC gap-filled products: Phase one methodology*. Retrieved from https://www.usgs.gov/media/files/landsat-7-slc-gap-filled-products-phase-one-methodology

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Preprint, *arXiv*:1409.1556.

Sundberg, R., & Melander, E. (2013). Introducing the UCDP georeferenced event dataset. *Journal of Peace Research*, *50*(4), 523–532.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA (pp. 1–9). Piscataway, NJ: IEEE.

USGS. (2017a). *Successful maneuver spells beginning of the end for Landsat 7*. Retrieved from https://www.usgs.gov/news/successful-maneuver-spells-beginning-end-landsat-7

USGS. (2017b). *USGS EarthExplorer*. Retrieved from https://earthexplorer.usgs.gov/

USGS. (2017c). *USGS Landsat Bulk Download*. Retrieved from https://ers.cr.usgs.gov/login>RET_ADDR=https%3A%2F%2Fearthexplorer.usgs.gov%2Fbulk%2F

USGS. (2018a). *Landsat 7*. Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-7

USGS. (2018b). *Landsat band specifications*. Retrieved from https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites

USGS. (2018c). *Landsat collections*. Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-collection-1

USGS. (2018d). *SLC-off products: Background*. Retrieved from https://www.usgs.gov/media/files/landsat-7-scan-line-corrector-processing-algorithm-theoretical-basis

USGS. (2019). *Landsat satellite missions*. Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-satellite-missions?qt-science_support_page_related_con=2#qt-science_support_page_related_con

Weiss, K. R., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*, 1–40.

Xie, S. M., Jean, N., Burke, M., Lobell, D. B., & Ermon, S. (2015). Transfer learning from deep features for remote sensing and poverty mapping. Preprint, *arXiv*:1510.00098.

Yang, L., Huang, C., Homer, C. G., Wylie, B. K., & Coan, M. J. (2003). An approach for mapping large-area impervious surfaces: Synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Canadian Journal of Remote Sensing*, *29*(2), 230–240.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. *27*, pp. 3320–3328). Red Hook, NY: Curran Associates.

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, CA (pp. 4559–4565). Menlo Park, CA: AAAI.

Zeng, C., Shen, H., & Zhang, L. (2013). Recovering missing pixels for Landsat ETM+ SLC-off imagery using multi-temporal regression analysis and a regularization method. *Remote Sensing of Environment*, *131*, 182–194.

Zhu, X., Liu, D., & Chen, J. (2012). A new geostatistical approach for filling gaps in Landsat ETM+ SLC-off images. *Remote Sensing of Environment*, *124*, 49–60.

How to cite this article: Goodman S, BenYishay A, Runfola D. A convolutional neural network approach to predict non-permissive environments from moderate-resolution imagery. *Transactions in GIS*. 2020;00: 1–18. https://doi.org/10.1111/tgis.12661